PROPOSAL

STRESS LEVEL PREDICTIVE MODEL

Joseph Choi | Jenny Overby | Daniel Meier | Serge Nane

DSC450 Applied Data Science
Spring 2024

DOMAIN

Our third and final project taps into the psychological and educational domain as we plan to develop a predictive model to forecast stress levels in students. Our goal is to study the stress factors affecting students today, diving deep into the psychological, physiological, social, environmental, and academic aspects of the topic.

We have listed ten online references that we believe will deepen our understanding of the domain and the data we'll use. These articles were chosen to get more detailed information regarding the correlation between the factors featured in our dataset and our target variable, stress levels. Through these texts, we are hoping to figure out the best approach to build out our model based on the insights we gather from these sources.

1. How stress-related factors affect mental wellbeing of university students A cross-sectional study to explore the associations between stressors, perceived stress, and mental wellbeing - PMC (nih.gov): Discusses a study that was done to understand the associations between stressors, perceived stress, and mental well-being of students attending universities.
2. Full article: The impact of stress on students in secondary school and higher education (tandfonline.com): Analyzes the impact of academic-related stress on students' mental health and how it affects academic performance.
3. Stress and Quality of Life Among University Students: A Systematic Literature Review - ScienceDirect: Provides information relating to the relationship between stress and quality of life among university students.
4. (PDF) Stress among students: An emerging issue (researchgate.net): Highlights the negative and positive impacts on the health and academic performance of students due to stress.
5. Full article: Academic stress as a predictor of mental health in university students (tandfonline.com): Dives into the relationship between academic stress and mental health among 1,265 university students
6. The Effects of Stress on College Students & Ways to Overcome it (bau.edu): Talks about the concept of academic stress and the psychological effects it has on college students.
7. Stress modelling and prediction in presence of scarce data - ScienceDirect: Discusses approaches for predicting stress levels using techniques like semi-supervised learning and ensemble methods.
8. Prediction of stress levels in the workplace using surrounding stress - ScienceDirect: Explores the prediction of stress levels by analyzing stress data. Achieved an 80% F-score using surrounding stress data.
9. Stress prediction using micro-EMA and machine learning during COVID-19 social isolation - ScienceDirect: Introduces a stress prediction system using micro-EMA historical data to forecast stress. Achieved absolute error of 4.26 and an accuracy of 81% in stress prediction.
10. Sensors | Free Full-Text | Stress Monitoring Using Machine Learning, IoT and Wearable Sensors (mdpi.com): Discusses a machine learning-based system that monitors stress by analyzing body temperature, sweat, and motion rate. Achieved an accuracy rate of 99.5%.

ABOUT THE DATASET

DESCRIPTION

For this project, the dataset we will be working with consists of factors influencing student stress levels, retrieved from a study examining the psychological, physiological, social, environmental, and academic elements of the topic. Sourced from Kaggle, we will be analyzing how these factors correlate with stress levels to develop an accurate predictive model for student stress. It is important to note that all features within this dataset are either scaled or rated.

DATA FIELDS

The Stress Level dataset contains 20 influential features that impact stress on students, grouped into 5 categories.

- PSYCHOLOGICAL FACTORS:
  - **anxiety_level**: Scale from 0 to 21 based on the GAD-7 scale
  - **self_esteem**: Scale from 0 to 30 using the Rosenberg Self-Esteem Scale
  - **mental_health_history**: Presence (1) or absence (0) of mental health history
  - **depression**: Level from 0 to 27, based on the Patient Health Questionnaire (PHQ-9)
- PHYSIOLOGICAL FACTORS:
  - **headache**: Rated from 0 (none) to 5 (high)
  - **blood_pressure**: Levels categorized from 0 (low) to 5 (high)
  - **sleep_quality**: Rated from 0 (poor) to 5 (excellent)
  - **breathing_problem**: Rated from 0 (none) to 5 (severe)
- ENVIRONMENTAL FACTORS:
  - **noise_level**: Rated from 0 (quiet) to 5 (loud)
  - **living_conditions**: Rated from 0 (poor) to 5 (excellent)
  - **safety**: Rated from 0 (unsafe) to 5 (very safe)
  - **basic_needs**: Rated from 0 (unmet) to 5 (fully met)
- ACADEMIC FACTORS:
  - **academic_performance**: Rated from 0 (poor) to 5 (excellent)
  - **study_load**: Rated from 0 (light) to 5 (heavy)
  - **teacher_student_relationship**: Rated from 0 (poor) to 5 (excellent)
  - **future_career_concerns**: Rated from 0 (none) to 5 (extreme)
- SOCIAL FACTORS:
  - **social_support**: Rated from 0 (none) to 5 (strong)
  - **peer_pressure**: Rated from 0 (none) to 5 (high)
  - **extracurricular_activities**: Rated from 0 (none) to 5 (high)
  - **bullying**: Rated from 0 (none) to 5 (frequent)
- **stress_level**: Rated from 0 (none) to 5 (extreme)

DATASET LOCATION

The dataset can be located and downloaded at the link: [Student Stress Factors: A Comprehensive Analysis (kaggle.com)]

RESEARCH QUESTION

In today's world, mental health is a major concern, especially for students. It has become evident that managing stress levels is essential as it can greatly affect one's life path and lead to negative outcomes if not handled properly. With students facing constant pressures of academic, social, and personal challenges, it is important to fully understand the factors influencing their stress levels and to what degree.

Our project will address the following research questions:

1. What are the key factors influencing stress levels among students?
2. Can we build a predictive model that can accurately predict stress levels based on these factors?

Insights gathered from this project will be key to maintaining the well-being of students and performing effectively in school. That is why our team would like to research this topic and get to the bottom of how best to support students and implement effective stress management strategies.

METHOD

The way we are approaching this project will be by going through the outlined process workflow below:

DATA WRANGLING

In this phase, our team will start by evaluating the quality of our data. Once we have a good grasp of our dataset and identify any discrepancies, we will begin cleaning and transforming the data to prepare it for the data science phase of the project.

DATA SCIENCE

During this phase, we will start with exploratory data analysis (EDA) to understand our input features and target variable in detail. We plan to rely on the information gathered from EDA to guide our feature selection, feature engineering, and model selection process. Once the dataset is ready, we will experiment with different models, document their performance, and select the best-performing model based on the evaluation metrics.

The initial plan is much like how we approached the data science phase in Project 2. However, this time, we are taking additional steps by researching the best methods for feature engineering and selecting the appropriate model based on our dataset's characteristics. Last time, our approach was more general and broad. This time, we would like to tailor it more closely to the specific characteristics of our dataset.

DATA VISUALIZATION

In this phase, we will build data visualizations that will support and address our research questions. In addition, we plan to create visuals for any important insights we extract during our analysis. These plots will help us communicate our findings in an easy-to-understand manner.

POTENTIAL ISSUES

In our last two projects, our model performances unfortunately fell short of where we were hoping they would be. We concluded that the low ratings were mainly due to our training approach and methods. When it comes down to it, this project's success depends on aligning tasks with our dataset's characteristics. If we do not, we could encounter the same results as before. Therefore, our goal is to be more intentional and logical in our approach this time. We'll work closely together to ensure a high-performing model by the end of the project.

CONCLUSION

Tying everything together, our project proposal focuses on understanding the influential factors that impact student stress levels. Through data wrangling, data analysis, data science, and data visualization, our team's primary objective is to identify key factors influencing stress and develop a predictive model that can accurately forecast stress levels. We are excited to get started and look forward to sharing our findings and the final model at the conclusion of the project.